



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/837,158	04/19/2001	Karen Mae Holland	ARC000018US1	8446
21254	7590	06/10/2004	EXAMINER	
MCGINN & GIBB, PLLC 8321 OLD COURTHOUSE ROAD SUITE 200 VIENNA, VA 22182-3817			BLACKWELL, JAMES H	
		ART UNIT	PAPER NUMBER	
		2176		

DATE MAILED: 06/10/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/837,158	HOLLAND ET AL.	
	Examiner	Art Unit	
	James H Blackwell	2176	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --
Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) Responsive to communication(s) filed on 19 April 2001.
 2a) This action is **FINAL**. 2b) This action is non-final.
 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4) Claim(s) 1-23 is/are pending in the application.
 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
 5) Claim(s) _____ is/are allowed.
 6) Claim(s) 1-23 is/are rejected.
 7) Claim(s) _____ is/are objected to.
 8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9) The specification is objected to by the Examiner.
 10) The drawing(s) filed on 19 April 2001 is/are: a) accepted or b) objected to by the Examiner.
 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
 Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
 a) All b) Some * c) None of:
 1. Certified copies of the priority documents have been received.
 2. Certified copies of the priority documents have been received in Application No. _____.
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892)	4) <input type="checkbox"/> Interview Summary (PTO-413)
2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)	Paper No(s)/Mail Date. _____
3) <input checked="" type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____	5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152)
	6) <input type="checkbox"/> Other: _____

DETAILED ACTION

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

Claims 1-23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Apte et al. (hereinafter Apte, U.S. Patent No. 6,654,739).

In regard to independent Claim 1 (and similarly independent Claim 23), Apte teaches a lightweight clustering method that uses a reduced indexing view of the original documents, where only the k best keywords of each document are indexed. An efficient procedure for clustering is specified in two parts: (a) compute k most similar documents for each document in the collection, and (b) group the documents into clusters using these similarity scores (Col. 3, lines 7-13; compare to Claim 1 (and similarly Claim 23), “*... generating a dictionary of keywords in said text documents; forming categories of said text documents using said dictionary and an automated algorithm*”). Apte does not specifically teach *counting occurrences of said structured variables, said categories and said structured variable/category combinations in said text documents*. However, Apte teaches reducing the number of terms or words for a given document by only indexing the top k-words (Col. 3, lines 50-65) which suggests that only the words or phrases used as keywords that occur most often are used, hence a predetermined number are used. Apte also does not

specifically teach calculating probabilities of occurrences of said structured variable/category combinations. However, Apte does suggest that probabilities are computed (Col. 7, lines 1-5). Apte does not specifically teach the concept of structured variables or of structured variable/category combinations. However, it would have been obvious to one of ordinary skill in the art at the time of invention to realize that structured variables, if they exist in a given document, would be treated in a similar fashion to the rest of the words in a document, and in a collection of documents as far as calculating probabilities of their occurrences in a single document or in a group of documents was concerned. The benefit would have been to determine the existence and statistical significance of words within documents.

In regard to dependent Claims 2-4, Claims 2-4 reflect the method for automatically identifying relationships between text documents and structured variables used for performing the methods as claimed in Claims 1 and 23 and are rejected along the same rationale.

In regard to dependent Claim 5, Apte does not specifically teach inputting a predetermined number of categories. However, Apte does teach reducing the number of terms or words for a given document by only indexing the top k-words (Col. 3, lines 50-65; compare with Claim 5, “... **said forming categories comprises inputting a predetermined number of categories**”).

In regard to dependent Claim 6 (and similarly dependent Claim 11), Apte does not specifically teach generating a sparse matrix array. However, Apte does discuss that a single document has a sparse vector over the complete set of words in all

documents (Col. 3, lines 57-58; compare with Claim 6 (and similarly Claim 11), “**... said forming categories comprises: generating a sparse matrix array containing a count of each of said keywords in each of said text documents**”).

In regard to dependent Claim 7, Apte teaches reducing the number of terms or words for a given document by only indexing the top k-words (Col. 3, lines 50-65) which suggests that only the words or phrases used as keywords that occur most often are used, hence a predetermined number are used. Compare with Claim 7, “**... said keywords comprise words or phrases which occur a predetermined number of times in said text documents**”).

In regard to dependent Claim 8, Apte does not specifically teach a Chi squared probability function. However, Apte does suggest that such probabilities are computed (Col. 7, lines 1-5; compare with Claim 8, “**... said calculating probabilities comprises using a Chi squared function**”).

In regard to dependent Claim 9, Apte teaches that the clustering algorithm of the present invention processes documents in a transformed state, where the documents are represented as a collection of terms or words. A vector representation is used. In the simplest format, each element of the vector is the presence or absence of a word. The same vector format is used for each document; the vector is a space taken over the complete set of words in all documents. Clearly, a single document has a sparse vector over the set of all words. Some processing may take place to stem words to their essential root and to transform the presence or absence of a word to a score, such as TF-TDF, that is a predictive distance measure. In addition, weakly predictive words (i.e.,

stop words) are removed. These same processes can be used to reduce indexing further by measuring for a document vector only the top k-words in a document and setting all remaining vector entries to zero (Col. 3, lines 50-65). Whether parsing for words or phrases, the technique is the same. Compare with Claim 9, “*... first parsing text in said text document to identify and count occurrences of words; storing a predetermined number of frequently occurring words; second parsing text in said text documents to identify and count occurrences of phrases; and storing a predetermined number of frequently occurring phrases*”).

In regard to dependent Claim 10, Apte does not teach *said frequently occurring words and phrases are stored in a hash table*. However, it would have been obvious to one of ordinary skill in the art at the time of invention to use a hash table as it was but one of several commonly used data structures used in association with clustering algorithms. The benefit would have been to provide a data structure for associating keywords with their number of occurrences.

In regard to dependent Claim 12, Apte does not specifically teach *said relationships comprise structured variable/category combinations having a lowest probability of occurrence*. However, it would have been obvious to one of ordinary skill in the art at the time of invention to realize that such a computation would have come from the Chi squared probability analysis taught in Claim 8. The benefit would have been to help determine whether or not a result was statistically significant.

In regard to dependent Claim 13, Apte teaches that a single clustering run, one row in Table 1 currently takes 15 minutes on a 375 MHz IBM RS6000 workstation

running AIX (IBM's version of the UNIX operating system). The code is written in the Java programming language (Col. 8, lines 30-33; compare with Claim 13, “***... said method comprises a computer implemented method***”).

In regard to dependent Claims 14 (and similarly dependent Claim 21), Claims 14 and 21 reflect the method of Claim 8 and are rejected along the same rationale.

In regard to dependent Claims 15-16 and 19-20, Apte does not teach that *structured variables comprise predetermined time intervals* or that the *predetermined time intervals comprise one of days, weeks, months and years*. However, structured variables defined in such a way would have been obvious to one of ordinary skill in the art at the time of invention because those variables could have been isolated as the result of the execution of a clustering scheme providing the benefit of comparing the statistical significance of the occurrence of selected words within a document or set of documents.

In regard to independent Claim 17, Apte teaches that a single clustering run, one row in Table 1 currently takes 15 minutes on a 375 MHz IBM RS6000 workstation running AIX (IBM's version of the UNIX operating system). The code is written in the Java programming language (Col. 8, lines 30-33). It is well known that a typical computer system consists of several input devices such as a keyboard and mouse, a processor for computing, memory for storing results, and a screen or other output device such as a printer for display purposes. Compare with Claim 17, “***... an input device for inputting text documents; a processor for forming categories of said text documents and counting occurrences of said structured variables,***

categories and structured variable/category combinations and calculating probabilities of occurrence of said structured variable/category combinations; and a display, for displaying said probabilities”.

In regard to dependent Claim 18, Claim 18 reflects the method of Claim 13 and is rejected along the same rationale.

In regard to dependent Claim 22, Claim 22 reflects the method of Claim 8 and is rejected along the same rationale.

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to James H Blackwell whose telephone number is 703-305-0940. The examiner can normally be reached on Mon-Fri.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Joseph H Feild can be reached on 703-305-9792. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

James H. Blackwell
06/03/04



JOSEPH FEILD
SUPERVISORY PATENT EXAMINER